WORLD JOURNAL OF PHARMACY AND BIOTECHNOLOGY
E-ISSN: 2349-9087 | https://doi.org/10.30904/j.wjpbt.2026.5037

Clinical Data Management in Pharmaceutical Research: Digital Transformation, Data Integrity, and Future Trends

Ravi Teja Gatla*[1], Muneendra Madam[2], Manjula Chella[3]

[1]*Quality assurance associate, ScieGen Pharmaceuticals, 89 Arkay Drive, Hauppauge, New York 11788, USA*
[2]*ScieGen Pharmaceuticals, 89 Arkay Drive, Hauppauge, New York 11788, USA*
[3]*Associate Professor, Department Pharmaceutical Technology, Vasavi Institute of Pharmaceutical Sciences, Vasavi Nagar, Peddapalli, Kadapa, A.P, India*

***Corresponding E-mail:** ravitejagatla1996@gmail.com*

KEYWORDS
Clinical trial,
CDM tools,
Data management,
Discrepancy management,
Regulatory audit,
Novel technologies

Received 30-04-2026
Revised 26-05-2026
Accepted 09-06-2026
Published 02-07-2026

ABSTRACT
Clinical trial is intended to find answers to the research question by means of generating data for proving or disproving a hypothesis. The quality of data generated plays an important role in the outcome of the study. CDM is the process of collection, cleaning, and management of subject data in compliance with regulatory standards. The primary objective of CDM processes is to provide high-quality data by keeping the number of errors and missing data as low as possible and gather maximum data for analysis. To meet better data quality best practices are adopted to ensure that data are complete, reliable, and processed correctly. Commonly used CDM tools are ORACLE CLINICAL, CLINTRIAL, MACRO, RAVE, and e-Clinical Suite. These CDM tools ensure the audit trail and help in the management of discrepancies. These software tools are expensive and need sophisticated Information Technology infrastructure to function. As a Pharmaceutical trial is designed to answer the research question, thereby CDM process will be eventually designed to deliver an error-free, valid, and statistically sound database. In regulatory submission studies, maintaining an audit trail of data management activities is of paramount importance. According to the roles and responsibilities, multiple user IDs can be created with access limitation to data entry, medical coding, database designing, or quality check. During a regulatory audit, the auditors can verify the discrepancy management process; the changes made and can confirm that no unauthorized or false changes were made. Developments on the technological front has positively impacted the CDM process and systems, thereby leading to encouraging results on speed and quality of data being generated.

INTRODUCTION

Clinical trials are an important research method for improving medical knowledge and patient care. Evidence has linked poor data quality to incorrect conclusions and recommendations; therefore, data quality is of paramount importance for acquiring reliable research findings from clinical trials. Many strategies and interventions have been developed aiming at reducing error in clinical trials, including standard operating procedures (SOPs), personnel training, data monitoring, and design of case report forms (CRFs). Additionally, multiple international and national guidelines stipulate the need for data quality and assurance; however, they are nonspecific in the nature and extent of the necessary methods [1-2].

According to ICHGCP, clinical trials must protect the rights and safety of all patients and ensure trial results are legible and valid. Onsite monitoring is important to achieve high data quality and to ensure the method of source data verification a (SDV) fulfills the original objectives. To optimize the efficacy of monitoring, there has been an emphasis on data audits and reducing onsite monitoring. However, the ICHGCP guidelines are flexible in interpretation and do not provide specific details on how and when to conduct audits, or how much or how little monitoring is required to maintain data integrity. There is a lack of evidence to support intensive monitoring for data quality; in fact, updated guidelines promote alternative monitoring methods, such as risk-based approaches b, centralized monitoring c and remote monitoring d, which complement trial procedures by improving the use of resources available.

It is evident that a standardized approach needs to be adopted for monitoring data quality in clinical research. In support of this notion, a survey conducted by the Clinical Trials Transformation Initiative investigated the intensity, focus, and methodology of monitoring practices by clinical research sponsors over a range of trial settings. As digital technology advances, health care organizations are confronted with even larger volume and more varied datasets produced by modern sensors and wearables, the adoption of electronic health records, and advancements in medical imaging and genome sequencing [3-5]. In addition, AI-driven workflows require not only large and varied datasets but also near real-time data processing to handle streaming information and provide timely insights. With

weak support of real-time data, these cDWHs delay detection of acute events reducing the window for early inventions. If data arrives in formats with inconsistent structures or with missing values, it can reduce the accuracy and reliability of AI outputs. To meet these increasing requirements, more recent concepts of data lakes and data lakehouses have been adapted to the clinical context, resulting in clinical data lakes (cDLs) and clinical data lakehouses (cDLHs).

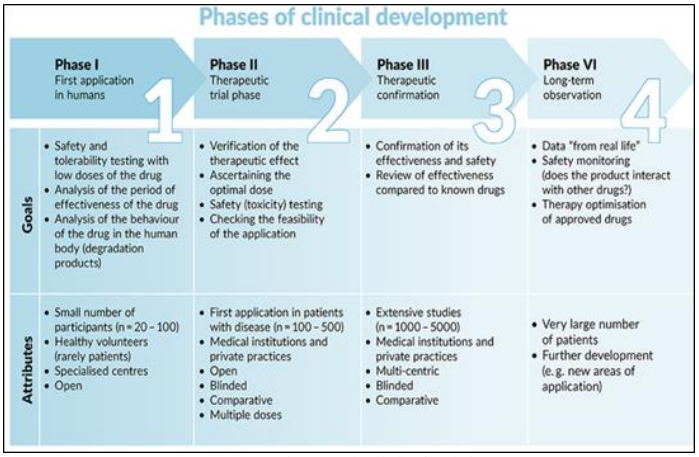
Phases of clinical development
Phase I
First application in humans
Phase II
Therapeutic trial phase
Phase III
Therapeutic confirmation
Phase VI
Long-term observation
1
2
3
4
Goals
• Safety and tolerability testing with low doses of the drug
• Analysis of the period of effectiveness of the drug
• Analysis of the behaviour of the drug in the human body (degradation products)
• Verification of the therapeutic effect
• Ascertaining the optimal dose
• Safety (toxicity) testing
• Checking the feasibility of the application
• Confirmation of its effectiveness and safety
• Review of effectiveness compared to known drugs
• Data "from real life"
• Safety monitoring (does the product interact with other drugs?)
• Therapy optimisation of approved drugs
Attributes
• Small number of participants (n = 20 – 100)
• Healthy volunteers (rarely patients)
• Specialised centres
• Open
• First application in patients with disease (n = 100 – 500)
• Medical institutions and private practices
• Open
• Blinded
• Comparative
• Multiple doses
• Extensive studies (n = 1000 – 5000)
• Medical institutions and private practices
• Multi-centric
• Blinded
• Comparative
• Very large number of patients
• Further development (e.g. new areas of application)

Fig.1: Phases of Clinical Trials

A cDL is an architectural approach that enables the storage of large amounts of raw data in its original format, including structured, semistructured, and unstructured data. This is in contrast to traditional cDWHs, which store data in a highly structured, organized way. cDLs are designed to provide high scalability and flexibility in data storage, allowing organizations to store data without the need for a defined schema first. This enables multimodal views on patients using diverse datasets, including real-time processing, but comes with drawbacks in data governance and data quality challenges for clinical decision-making.

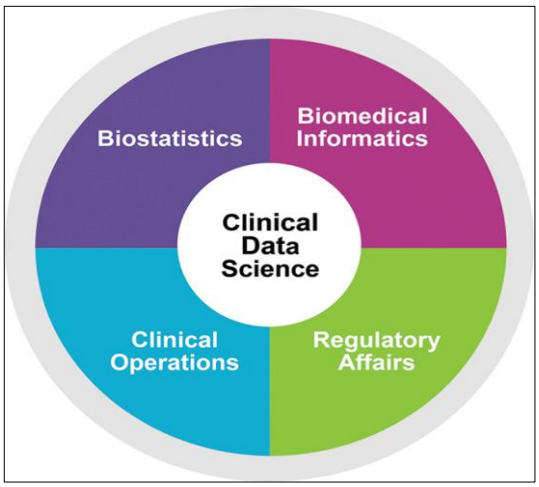

Figure 2: Theoretical framework for a clinical data science

GCP Guidelines

- All clinical research data should be recorded, handled, & stored in a way that allows its accurate reporting, interpretation & verification. (ICH GCP 2.10, 4.9, 5.5, 5.14 & ICH E9 3.6 & 5.8)
- Systems with procedures that assure quality of every aspect of research should be implemented. (GCP 2.13).
- Quality assurance & quality control systems with written standard operating procedures should be implemented & maintained to ensure that research are conducted & data are generated, documented & recorded, & reported in compliance with protocol, GCP & applicable regulatory requirements. (GCP 5.1.1).
- If data are transformed during processing, it should always be possible to compare original data & observations with processed data (ICH GCP 5.5.4).
- Sponsor should use an unambiguous subject identification number or code that allows identification of all data reported for each subject. (ICH GCP 5.5.5)
- Protocol amendments that necessitate a change in design of CRF, subject diaries, study worksheets, research database & other key aspects of CDM processes need to be controlled. (ICH E9 2.1.2).
- Common standards should be adopted for a number of features of research such as dictionaries of medical terms, definition & timing of main measurements, handling of protocol deviations. (ICH E9 2.1.1).

Regulatory Bodies in Clinical Data Management

- FDA (Food and Drug Administration): Overview of FDA's role in regulating clinical trials and ensuring data integrity.
- EMA (European Medicines Agency): EMA's requirements for data management and quality in clinical research within the European Union.
- ICH (International Council for Harmonisation): Importance of ICH guidelines (e.g., E6(R2) for GCP) in establishing international standards.

Clinical Data Management in Pharmaceutical research

Clinical trial is intended to find answers to the research question by means of generating data for proving or disproving a hypothesis. The quality of data generated plays an important role in the outcome of the study. CDM is the process of collection, cleaning, and management of subject data in compliance with regulatory standards. The primary objective of CDM processes is to provide high-quality data by keeping the number of errors and missing data as low as possible and gather maximum data for analysis. To meet this objective, best practices are adopted to ensure that data are complete, reliable, and processed correctly. This has been facilitated by the use of software applications that maintain an audit trail and provide easy identification and resolution of data discrepancies. Sophisticated innovations have enabled CDM to handle large trials and ensure the data quality even in complex trials.

Objective of clinical data management

The primary objective of clinical data management is to provide high-quality, reliable data for reporting randomized controlled trials in line with good clinical practice requirements. For treatment trials of neglected tropical diseases in endemic countries, CDM systems need to be efficient and affordable. Challenges include poor infrastructure, license costs associated with GCP-compliant software, and limited human resources to provide the required expertise. We argue that high-quality CDM for NTDs can be achieved and that challenges can be overcome through the use of readily available open-access tools [6-8].

Use of Open-Source Tools for CDM in Remote Settings

A popular tool for RCTs is OpenClinica, a GCP-compliant database package used in our DC since 2009. OpenClinica Community Edition has many advantages for CDM but was only available to use in an online mode. Without a validated offline mode of working with OpenClinica, the system's ability to support multicenter trials is limited to paper data collection and centralized data entry. Offline electronic data capture (EDC) in line with GCP would lead to substantial reductions in time between data entry and database lock.

Data synchronization was enabled by developing codes using python scripts, coupled with dedicated technical support and user training. As a result, actual entry of data into the system in a remote setting is faster because the delays associated with poor connectivity are eliminated and the time taken for the data to reach the DC is shortened due to the fact that paper case report forms (CRFs) do not require physical transportation, site capacity has been strengthened, and the site staff feel engaged with the project beyond patient management through participation in CDM.

Tools for CDM

Many software tools are available for data management, and these are called Clinical Data Management Systems (CDMS). In multicentric trials, a CDMS has become essential to handle the huge amount of data. Most of the CDMS used in pharmaceutical companies are commercial, but a few open source tools are available as well. Commonly used CDM tools are ORACLE CLINICAL, CLINTRIAL, MACRO, RAVE, and eClinical Suite. In terms of functionality, these software tools are more or less similar and there is no significant advantage of one system over the other. These software tools are expensive and need sophisticated Information Technology infrastructure to function. Additionally, some multinational pharmaceutical giants use custom-made CDMS tools to suit their operational needs and procedures. Among the open source tools, the most prominent ones are OpenClinica, openCDMS, TrialDB, and PhOSCo. These CDM software are available free of cost and are as good as their commercial counterparts in terms of functionality [9-12].

Table 1: Online resources for Clinical data management

Resource	Link
Data Management Tools:	
OpenClinica: open-source database software.	www.openclinica.com
OpenClinica offline module: a user-written module to allow data entry without internet connectivity.	http://en.wikibooks.org/wiki/OpenClinica_User_Manual/OfflineDistributions
	http://blog.openclinica.com/2013/07/17/another-option-for-using-openclinica-in-disconnected-settings/
R: R is a language and environment for statistical computing and graphics and is available as free software under the terms of the Free Software Foundation's GNU General Public License in source code form.	www.r-project.org
Regulatory:	
US Food and Drug Administration Guidelines for Computerized Systems Used in Clinical Investigations.	www.fda.gov/OHRMS/DOCKETS/98fr/04d-0440-gdl0002.PDF
International Conference of Harmonization Good Clinical Practice (ICH GCP): GCP for Data Management.	International Conference on Harmonisation. *ICH Expert Working Group: Guideline for Good Clinical Practice, E6*, Section 5.1.3; 1996.
Networks:	
Society of Clinical Data Managers (SCDM).	www.scdm.org/
Global Health Trials.	http://globalhealthtrials.tghn.org
Association for Data Management in the Tropics(ADMIT).	http://admit.tghn.org

Types of Clinical Data Management

Clinical Data Management (CDM) deals specifically with the information that is collected during the clinical trial. Its aims are to ensure quality, data, integrity and compliance with regulations and protocols.

Data Management Stages

There are several areas of responsibility when it comes to data management. These responsibilities can either be split into smaller groups or assigned to the principal data manager. These roles include [13-15]:

Planning: The database, forms and overall plans are prepared by the Data Manager

Data collection and management: Data collection, and conversion to digital (if required), coding and standardization.

Assure: The Data Manager determines if the data plan and tools meet the necessary requirements

Identity: Any risks are identified by the staff

Data review and analytics: Quality management, auditing and statistical analysis of the collected information.

Data standards: Comparing the data collected against regulatory requirements.

Innovation: Use new tools and theories that are developed around the field of medical research and clinical trials.

Lock: The database is locked for integrity by the Data Manager.

Different Types of Data Management

Data Management Plan Design

A Data Management Plan is a document that outlines all the tasks, processes, milestones and deliverables that are required throughout the Clinical Data Management process. It is a clear guideline on how to work with the information gathered during the clinical trial and how to handle any possible risks. The Data Manager and Database Designer are usually in charge of the DMP. The DMP must be completed during the trial design stage [16-18].

The DMP involves the following aspects:

- Data formats
- Any existing information that can be integrated
- Storage and backup
- Data quality processes
- Responsibilities of each team member

- Access
- Archiving
- Cost of data preparation
- Compliance with relevant laws and regulations

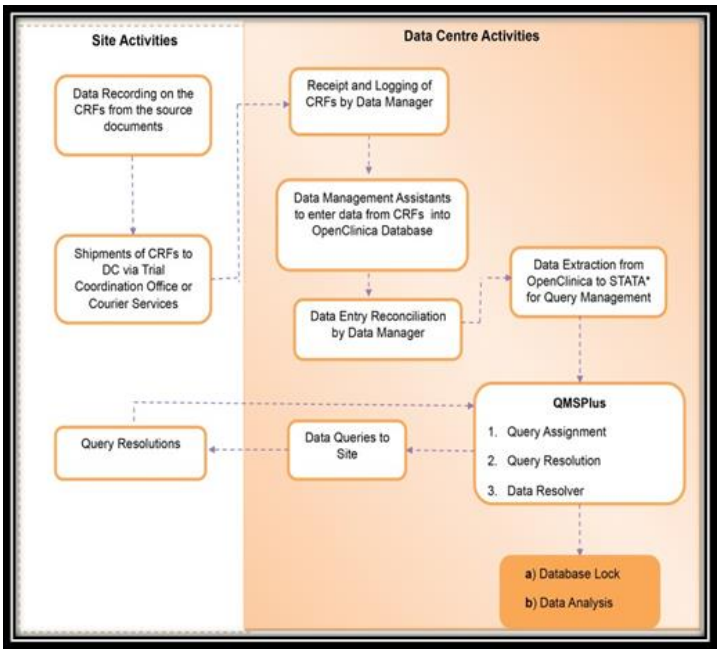
Site Activities
Data Centre Activities
Data Recording on the CRFs from the source documents
Receipt and Logging of CRFs by Data Manager
Data Management Assistants to enter data from CRFs into OpenClinica Database
Shipments of CRFs to DC via Trial Coordination Office or Courier Services
Data Extraction from OpenClinica to STATA* for Query Management
Data Entry Reconciliation by Data Manager
QMSPlus
1. Query Assignment
2. Query Resolution
3. Data Resolver
Query Resolutions
Data Queries to Site
a) Database Lock
b) Data Analysis

Figure 3: Process flow for query management

Electronic Case Report Form Design (e-CRF)

The Case Report Form is a digital or printed survey that collects important data from clinical trial participants and is shared with the trial sponsors. The document is created in accordance with the trial protocol and the recommendations of the Clinical Data Acquisitions Standards Harmonization. The Data Manager and Database Designer are responsible for the Case Report Form. There are several reviews of the design before one is finalized.

The information included in the Case Report Form includes:

- Demographics (gender and age)
- Basic measurements (height and weight)
- Vital signs (temperature and blood pressure, etc)
- Medical history
- Lab exams
- Any adverse events

Electronic Data Capture

In recent years a digital system was created to collect participant data in an electronic format. More Clinical Research Organization's (CROs) and pharmaceutical companies are turning towards EDCs as it is usually user-friendly and minimizes any errors [19-20].

The data is collected from:

- Electronic Health Records (EHRs)
- Medical devices (blood pressure monitors and ECG machines, etc)
- Patient-reported outcomes (PROs) or information that comes directly from the participant
- Clinicians, Data Managers, Data Entry Associates and Medical Coders are in charge of Electronic Data Capture.

Traditional Data Capture Methods

Traditional data capture is comprised of data being manually entered into a database in a consistent manner so that it can easily be retrieved and searched. Data entries may consist of a single entry or double entries which increases the accuracy of the data.

Data Entry Screen Design

The design of the data entry screen determines the speed at which data can be entered. The more similar the screen looks to the original CRF page, the easier it is for data to be entered into the system efficiently.

Integrating Data From an External Database

More companies are choosing to connect to external databases to simplify the transfer of large amounts of data to their own database. Data files can be copied onto a disc and uploaded directly to the sponsor database. Ensuring that the two databases are compatible and validating both systems to maintain data integrity and security is crucial [21-22].

Fax-Based Data Capture

The use of faxes to transfer CRFs is a good middle ground between more traditional paper-based data capture and full electronic data capture. It requires a simple re-engineering of the paper CRF process.

Remote Data Entry (RDE)

Capturing data at the point at which it is generated. If an electronic transfer of the data to the sponsor's system occurs on a regular basis, it means that it can be accessed and reviewed more quickly. The data is validated as it is entered into the system which minimizes any errors. The pros of using remote data entry include better data quality, speed, and access to data. The potential users of the system should be considered before investing in a remote data entry system. Once the data has been uploaded to the system, sponsors and CROs can access the information immediately. The edit history can be checked from the date stamps and once all edits and checks have been completed, the eCRF can be printed out and returned to the site by the CRA where it is signed by the investigator after the final data verification check. RDE is most effective for simple clinical trials with a large number of participants recruited at a limited number of sites.

Electronic Diaries

Data can be collected remotely from electronic diaries that have been given to participants. These diaries can be programmed to remind participants to follow the treatment schedule and regularly reviews their compliance. It also allows for adverse reactions to be picked up faster. Participants will need to be trained on how to use these diaries and will also need to remember to bring them with them on each visit [23].

Data Validation: Data Validation is a group of quality tests that ensure the accuracy, legibility, consistency and integrity of information. The Data Manager, Database Designer and Quality Control Associate are responsible for data validation.

It includes the following steps:

Electronic edit checks: Established by a Database Designer and embedded into Electronic Case Report Forms which compare inputs against numerical and logical data.

Source data verification: A process of checking Case Report Forms against medical records and other records. This ensures that the e-CRF contains all the relevant information and a participant's profile is fleshed out.

Data anonymization: Before it's submitted to the clinical trial sponsors, the data must be made anonymous.

Database Lock and Data Archiving

Once a clinical trial has been completed, the database needs to be locked so that no edits can be made. The data is then sent to stakeholders for analysis, reporting and publication of the results. All documents have to be archived for at least three years to allow for post-trial monitoring and trial follow-up to facilitate further research. The Data Manager and Database Designer are in charge of locking the database and archiving the information [24].

CDM Process

The CDM process, like a clinical trial, begins with the end in mind. This means that the whole process is designed keeping the deliverable in view. As a clinical trial is designed to answer the research question, the CDM process is designed to deliver an error-free, valid, and statistically sound database. To meet this objective, the CDM process starts early, even before the finalization of the study protocol.

Regulations, Guidelines in CDM

Akin to other areas in clinical research, CDM has guidelines and standards that must be followed. Since the pharmaceutical industry relies on the electronically captured data for the evaluation of medicines, there is a need to follow good practices in CDM and maintain standards in electronic data capture. These electronic records have to comply with a Code of Federal Regulations (CFR), 21 CFR Part 11. This regulation is applicable to records in electronic format that are created, modified, maintained, archived, retrieved, or transmitted. This demands the use of validated systems to ensure accuracy, reliability, and consistency of data with the use of secure, computer-generated, time-stamped audit trails to independently record the date and time of operator entries and actions that create, modify, or delete electronic records [25].

Society for Clinical Data Management (SCDM) publishes the Good Clinical Data Management Practices (GCDMP) guidelines, a document providing the standards of good practice within CDM. GCDMP was initially published in September 2000 and has undergone several revisions thereafter. The July 2009 version is the currently followed GCDMP document. GCDMP provides guidance on the accepted practices in CDM that are consistent with regulatory practices. Clinical Data Interchange Standards Consortium (CDISC), a multidisciplinary non-profit organization, has developed standards to support acquisition, exchange, submission, and archival of clinical research data and metadata. Metadata is the data of the data entered. This includes data about the individual who made the entry or a change in the clinical data, the date and time of entry/change and details of the changes that have been made. Among the standards, two important ones are the Study Data Tabulation Model Implementation Guide for Human Clinical Trials (SDTMIG) and the Clinical Data Acquisition Standards Harmonization (CDASH) standards, available free of cost from the CDISC website. The SDTMIG standard describes the details of model and standard terminologies for the data and serves as a guide to the organization [26].

Review and finalization of study documents

The protocol is reviewed from a database designing perspective, for clarity and consistency. During this review, the CDM personnel will identify the data items to be collected and the frequency of collection with respect to the visit schedule. A Case Report Form (CRF) is designed by the CDM team, as this is the first step in translating the protocol-specific activities into data being generated. The data fields should be clearly defined and be consistent throughout. The type of data to be entered should be evident from the CRF.

Similarly, the units in which measurements have to be made should also be mentioned next to the data field. The CRF should be concise, self-explanatory, and user-friendly (unless you are the one entering data into the CRF). Along with the CRF, the filling instructions (called CRF Completion Guidelines) should also be provided to study investigators for error-free data acquisition. CRF annotation is done wherein the variable is named according to the SDTMIG or the conventions followed internally. Annotations are coded terms used in CDM tools to indicate the variables in the study.

Database designing

Databases are the clinical software applications, which are built to facilitate the CDM tasks to carry out multiple studies. Generally, these tools have built-in compliance with regulatory requirements and are easy to use. "System validation" is conducted to ensure data security, during which system specifications, user requirements, and regulatory compliance are evaluated before implementation. Study details like objectives, intervals, visits, investigators, sites, and patients are defined in the database and CRF layouts are designed for data entry. These entry screens are tested with dummy data before moving them to the real data capture.

Data collection

Data collection is done using the CRF that may exist in the form of a paper or an electronic version. The traditional method is to employ paper CRFs to collect the data responses, which are translated to the database by means of data entry done in-house. These paper CRFs are filled up by the investigator according to the completion guidelines. In the e-CRF-based CDM, the investigator or a designee will be logging into the CDM system and entering the data directly at the site. In e-CRF method, chances of errors are less, and the resolution of discrepancies happens faster.

CRF tracking

The entries made in the CRF will be monitored by the Clinical Research Associate (CRA) for completeness and filled up CRFs are retrieved and handed over to the CDM team. The CDM team will track the retrieved CRFs and maintain their record. CRFs are tracked for missing pages and illegible data manually to assure that the data are not lost. In case of missing or illegible data, a clarification is obtained from the investigator and the issue is resolved.

Data entry

Data entry takes place according to the guidelines prepared along with the DMP. This is applicable only in the case of paper CRF retrieved from the sites. Usually, double data entry is performed wherein the data is entered by two operators separately. The second pass entry (entry made by the second person) helps in verification and reconciliation by identifying the transcription errors and discrepancies caused by illegible data. Moreover, double data entry helps in getting a cleaner database compared to a single data entry. Earlier studies have

shown that double data entry ensures better consistency with paper CRF as denoted by a lesser error rate.

Data validation

Data validation is the process of testing the validity of data in accordance with the protocol specifications. Edit check programs are written to identify the discrepancies in the entered data, which are embedded in the database, to ensure data validity. These programs are written according to the logic condition mentioned in the DVP. These edit check programs are initially tested with dummy data containing discrepancies. Discrepancy is defined as a data point that fails to pass a validation check. Discrepancy may be due to inconsistent data, missing data, range checks, and deviations from the protocol. In e-CRF based studies, data validation process will be run frequently for identifying discrepancies. These discrepancies will be resolved by investigators after logging into the system. Ongoing quality control of data processing is undertaken at regular intervals during the course of CDM. If for any patient, the condition becomes TRUE, a discrepancy will be generated. These discrepancies will be highlighted in the system and Data Clarification Forms (DCFs) can be generated [27].

Discrepancy management

This is also called query resolution. Discrepancy management includes reviewing discrepancies, investigating the reason, and resolving them with documentary proof or declaring them as irresolvable. Discrepancy management helps in cleaning the data and gathers enough evidence for the deviations observed in data. Almost all CDMS have a discrepancy database where all discrepancies will be recorded and stored with audit trail.

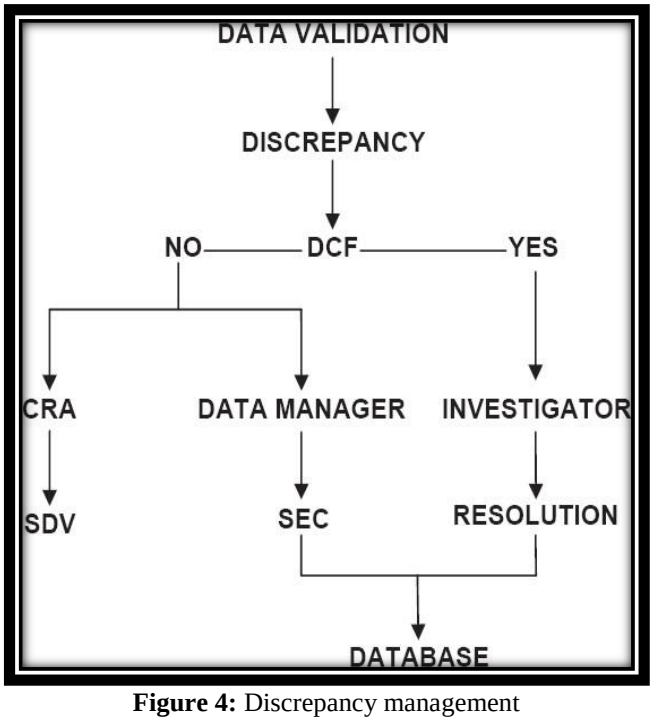

Figure 4: Discrepancy management

Medical coding

Medical coding helps in identifying and properly classifying the medical terminologies associated with the clinical trial. For classification of events, medical dictionaries available online are used. Functionally, it also requires knowledge about the structure of electronic medical dictionaries and the hierarchy of classifications available in them. Commonly, Medical Dictionary for Regulatory Activities (MedDRA) is used for the coding of adverse events as well as other illnesses and World Health Organization–Drug Dictionary Enhanced (WHODDE) is used for coding the medications. Medical coding helps in classifying reported medical terms on the CRF to standard dictionary terms in order to achieve data consistency and avoid unnecessary duplication. The right coding and classification of adverse events and medication is crucial as an incorrect coding may lead to masking of safety issues or highlight the wrong safety concerns related to the drug.

Roles and Responsibilities in Clinical Data Management

In a CDM team, different roles and responsibilities are attributed to the team members. The minimum educational requirement for a team member in CDM should be graduation in life science and knowledge of computer applications. Ideally, medical coders should be medical graduates. However, in the industry, paramedical graduates are also recruited as medical coders. Some key roles are essential to all CDM teams [28-29]. The list of roles given below can be considered as minimum requirements for a CDM team:

- Data Manager
- Database Programmer/Designer
- Medical Coder
- Clinical Data Coordinator
- Quality Control Associate
- Data Entry Associate

Need of Clinical Data Management

Clinical research studies provide crucial information. For example, results should demonstrate that a product is safe and meets requirements [30-32]. Clinical data management provides:

- Assurance of data quality
- Accelerated development
- Protection from data loss
- Reduced expenses
- Security
- Complete and accurate collection of data
- A clean dataset to support statistical analysis and reporting
- A formatted dataset for optimal and timely usability
- Assurance of data integrity and quality during database transfer
- True representation of the trial in the study database

Research methodologies for healthcare transformation

The emerging science of healthcare delivery (HCDS) has begun addressing this complex system problem, but many challenges and unanswered questions remain. In the opinion of Dr. James Weinstein, CEO and President of Dartmouth-Hitchcock Medical Center, healthcare delivery science "is the advance that is missing from healthcare reform legislation and without it will fail". One way of conceptualizing HCDS is to think about it as working on the healthcare delivery system as opposed to working in the healthcare system [33].

Digital transformation

Digital transformation refers to the digital technology changes used to benefit society and the healthcare industry. Healthcare systems need to use digital technology for innovative solutions to improve healthcare delivery and to achieve improvement in

medical problems. The digital transformation of healthcare includes changes related to the internet, digital technologies, and their relation to new therapies and best practices for better health management procedures. The quality control of massive data collected can help improve patients' well-being and reduce the cost of services. Digital technologies will also impact medical education, and experts will deceive new ways to train people. Now in this way, practitioners will face new opportunities.

Digital transformation is an ongoing process that can create opportunities in the health sector, provided the necessary infrastructure and training are available. Under Regulation (EU) 2021/694 of the European Parliament and of the Council of 29 April 2021, establishing the Digital Europe Program and repealing Decision (EU) 2015/2240, digital transformation is defined as the use of digital technologies for the transformation of businesses and services. Some technologies that contribute to digital transformation are the digital platform of the Internet of Things, cloud computing and artificial intelligence. At the same time, the sectors of society that are almost affected are telecommunications, financial services and healthcare. Digital health can play a role in innovation in health, as it facilitates the participation of patients in the process of providing health care [34].

Future trends of Clinical Data Management:

AI and Machine Learning (ML) Automation: Artificial intelligence is shifting from a novel tool to a standard component of everyday data practice. AI models automate data cleaning, validate entries in real time, and use natural language processing to parse unstructured text from electronic health records [35].

Real-World Evidence Integration: There is a strong industry shift toward patient-centric trials that incorporate data from everyday settings. Platforms now routinely aggregate data from wearable devices, mobile applications, and direct patient-reported outcomes to provide a more comprehensive view of patient health.

Risk-Based Quality Management: Rather than performing 100% source data verification, CDM teams rely on centralized monitoring strategies to identify and mitigate anomalies in real time. This aligns with modern regulatory frameworks like ICH GCP E6 (R3), ensuring rapid issue resolution.

Decentralized and Hybrid Trial Architectures: The operational landscape is shifting to support trial execution at diverse locations, requiring clinical platforms to seamlessly handle continuous streams of incoming data.

Stricter Data Compliance and Security: As the volume of patient data increases, organizations must navigate strict, evolving global data protection standards (such as GDPR and HIPAA).

CONCLUSION

Clinical data management contains diverse team of specialized professionals, each playing a crucial role in ensuring the success and integrity of clinical trials. It demonstrates data acquisition, data review, data cleaning and preparation for data analysis, and submission. CDM has more demand from pharmaceutical companies to accelerate the drug development process to ensure generation of high-quality data for accurate drug evaluation. To meet this required goals there is a gradual shift from the paper-based to the electronic systems of data management. Novel developments on the technological front has positively impacted the CDM process and systems, thereby leading to encouraging results on speed and quality of data.

From the industry side the biggest challenge would be the planning and implementation of data management systems in a changing operational environment where the rapid pace of technology development is outdating is being present with existing infrastructure. However, CDM is evolving to become a standard-based clinical research entity, achieved by maintaining the balance between technological developments and business demands. Implementing advanced analytics tools enables teams to turn complex data sets into meaningful insights with the potential to greatly improve patient outcomes. Effective clinical data management assist to analyze the number of errors and amount of missing data to a minimum. High-quality data enables teams, researchers, and organizations to garner insights and draw important analyze of data with document audit trails.

CONFLICT OF INTERESTS

The authors declare no conflict of interest

ETHICS APPROVAL

Not applicable

FUNDING: This study received no specific funding from public, commercial, or not-for-profit funding agencies.

AI TOOL DECLARATION

The authors declare that no AI and related tools are used to write the scientific content of this manuscript.

DATA AVAILABILITY

Data will be available on request

REFERENCES

[1] Bellazzi R. Big data and biomedical informatics: a challenging opportunity. Yearb Med Inform. 2014 May 22;9(1):8–13.

[2] Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract. 2017; 36(1):3–11.

[3] Ting DS, Deshmukh R, Ting DS, Ang M. Big data in corneal diseases and cataract: current applications and future directions. Front Big Data. 2023; 6:1017420.

[4] Natarajan K, Weng C, Sengupta S. A model for multi-institutional clinical data repository. Stud Health Technol Inform. 2023 May 18; 302:312–6.

[5] Lacey Jr JV, Chung NT, Hughes P, Benbow JL, Duffy C, Savage KE, Spielfogel ES, Wang SS, Martinez ME, Chandra S. Insights from adopting a data commons approach for large-scale observational cohort studies: the California teachers study. Cancer Epidemiol Biomarkers Prev. 2020 Apr; 29(4):777–86.

[6] Guilbert B, Rieu T, Delamarre D, Laurent F, Serre F, Leroy S, Paillard E, Le Buhan N, Ribault A, Cuggia M. ONCO-FAIR project: improving data interoperability in oncology chemotherapy treatments for data reuse.

Stud Health Technol Inform. 2024 Aug 22; 316:1373–7.

[7] AL-Hgaish A, ALzyadat W, Al-Fayoumi M, Alhroob AM, Thunibat A. Preserve quality medical drug data toward meaningful data lake by cluster. Int J Eng Technol. 2019 Sep 30;8(3):270–7.

[8] Dhayne H, Haque R, Kilany R, Taher Y. In search of big medical data integration solutions - a comprehensive survey. IEEE Access. 2019;7:91265–90.

[9] Nambiar A, Mundra D. An overview of data warehouse and data lake in modern enterprise data management. Big Data Cogn Comput. 2022 Nov 07;6(4):132.

[10] Quix C, Hai R. Data lake. In: Sakr S, Zomaya AY, editors. Encyclopedia of Big Data Technologies. Cham, Switzerland: Springer; 2018. pp. 1–8.

[11] Gentner T, Neitzel T, Schulze J, Gerschner F, Theissler A. Data lakes in healthcare: applications and benefits from the perspective of data sources and players. Procedia Comput Sci. 2023;225:1302–11.

[12] Lu Z, Su J. Clinical data management: Current status, challenges, and future directions from industry perspectives. Open Access J Clin Trials. 2010;2:93–105.

[13] Fegan GW, Lang TA. Could an open-source clinical trial data-management system be what we have all been looking for? PLoS Med. 2008;5:e6.

[14] Kuchinke W, Ohmann C, Yang Q, Salas N, Lauritsen J, Gueyffier F, et al. Heterogeneity prevails: The state of clinical trial data management in Europe - results of a survey of ECRIN centres. Trials. 2010;11:79.

[15] Reynolds-Haertle RA, McBride R. Single vs. double data entry in CAST. Control Clin Trials. 1992;13:487–94.

[16] Ottevanger PB, Therasse P, van de Velde C, Bernier J, van Krieken H, Grol R, et al. Quality assurance in clinical trials. Crit Rev Oncol Hematol. 2003;47:213–35.

[17] Haux R, Knaup P, Leiner F. On educating about medical data management - the other side of the electronic health record. Methods Inf Med. 2007;46:74–9.

[18] Bakobaki J M, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. Clin Trials. 2012;9(02):257–264.

[19] De S. Hybrid approaches to clinical trial monitoring: practical alternatives to 100% source data verification. Perspect Clin Res. 2011;2(03):100–104.

[20] Macefield R C, Beswick A D, Blazeby J M, Lane J A. A systematic review of on-site monitoring methods for health-care randomised controlled trials. Clin Trials. 2013;10(01):104–124.

[21] Brosteanu O, Schwarz G, Houben P et al. Risk-adapted monitoring is not inferior to extensive on-site monitoring: results of the ADAMON cluster-randomised study. Clin Trials. 2017; 14(06): 584–596.

[22] Tantsyura V, Dunn I M, Waters J et al. Extended risk-based monitoring model, on-demand query-driven source data verification, and their economic impact on clinical trial operations. Ther Innov Regul Sci. 2015;50(01):115–122.

[23] Olsen R, Bihlet A R, Kalakou F, Andersen J R. The impact of clinical trial monitoring approaches on data integrity and cost–a review of current literature. Eur J Clin Pharmacol. 2016;72(04):399–412.

[24] Mealer M, Kittelson J, Thompson B T et al. Remote source document verification in two national clinical trials networks: a pilot study. PLoS One. 2013; 8(12): e81890.

[25] Tudur Smith C, Stocken D D, Dunn J et al. The value of source data verification in a cancer clinical trial. PLoS One. 2012;7(12):e51623.

[26] Journot V, Pignon J P, Gaultier C et al. Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies–the Pre-Optimon study. Contemp Clin Trials. 2011;32(01):16–24.

[27] Morrison B W, Cochran C J, White J G et al. Monitoring the quality of conduct of clinical trials: a survey of current practices. Clin Trials. 2011; 8(03): 342–349.

[28] Krishnankutty B, Bellary S, Kumar N B, Moodahadu L S. Data management in clinical research: an overview. Indian J Pharmacol. 2012;44(02):168–172.

[29] Califf R M, Karnash S L, Woodlief L H. Developing systems for cost-effective auditing of clinical trials. Control Clin Trials. 1997;18(06):651-660.

[30] Eisenstein E L, Lemons P W, II, Tardiff B E, Schulman K A, Jolly M K, Califf R M. Reducing the costs of phase III cardiovascular clinical trials. Am Heart J. 2005; 149(03):482–488.

[31] Andersen J R, Byrjalsen I, Bihlet A et al. Impact of source data verification on data quality in clinical trials: an empirical post hoc analysis of three phase 3 randomized clinical trials. Br J Clin Pharmacol. 2015; 79(04):660-668.

[32] Kleppinger C F, Ball L K. Building quality in clinical trials with use of a quality systems approach. Clin Infect Dis. 2010;51 01:S111–S116.

[33] Arts D GT, De Keizer N F, Scheffer G J. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002;9(06):600-611.

[34] Rostami R, Nahm M, Pieper C F. What can we learn from a decade of database audits? The Duke Clinical Research Institute experience, 1997-2006. Clin Trials. 2009;6(02):141–150.

[35] Hullsiek K H, Kagan J M, Engen N et al. Investigating the efficacy of clinical trial monitoring strategies: design and implementation of the Cluster Randomized START Monitoring Substudy. Ther Innov Regul Sci. 2015; 49(02):225–233.